

与大数据同行的学习与教育

——《大数据时代》作者舍恩伯格教授和库克耶先生访谈

赵中建 张燕南

摘要 访谈问题主要集中于如下若干方面,即大数据与学校教育系统的重塑、大数据关照下的数字鸿沟问题、大数据时代背景下的教师与学校管理者、大数据与求变且渐变的学校教育模式、大数据的潜在威胁与可能的应对策略以及大数据的背后其实是人的问题等。舍恩伯格和库克耶对这些问题作了较为详尽的回答。

关键词 大数据; 学习; 慕课; 隐私

作者简介 赵中建/华东师范大学课程与教学研究所教授 (上海 200062)

张燕南/华东师范大学课程与教学研究所博士研究生 (上海 200062)

维克托·迈尔-舍恩伯格(Viktor Mayer-Schönberger)系牛津大学互联网研究所(Oxford Internet Institute)教授,作为大数据研究的权威人士,舍恩伯格共撰写了包括《大数据时代:生活、工作与思维的大变革》和《删除:大数据取舍之道》等在内的上百篇论文和八本著作。肯尼思·库克耶(Kenneth Cukier)现为《经济学人》(the Economist)杂志数据编辑,同时也是一位著名的大数据发展评论员。

舍恩伯格和库克耶于2014年初出版了《与大数据同行——学习与教育的未来》(Learning with BIG DATA: The Future of Education)一书,畅谈大数据与学习和教育的关系以及大数据将如何深刻影响到学习与教育的未来态势。我们在翻译《与大数据同行》一书的基础上,在上海会见了舍恩伯格教授并在当面讨论大数据及与大数据密切相关的教育技术革命等诸多话题的前提下,就书中的若干话题求教于舍恩伯格和库克耶,恳请他们给予进一步的解答和阐释。现将这些问答以访谈形式刊登于此,以飨读者。

一、大数据与学校教育系统的重塑

赵和张:企业可以利用大数据调整产品与服务,更好地满足消费者的需求;优化运营和基础设施建设;寻求更多资源支持营业额的增长(如IBM)。对教育领域来说,利用大数据的目的除了满足教师和学生的需求,提供教育公平机会之外,还可以在哪些方面也带来益处?

舍恩伯格和库克耶:在教育领域中,由于学习材料和学习工具正愈益贴近学生个体的需求和偏好,因此大数据能够有助于学生学得更为有效。大数据还能

帮助教师针对特定的学生群体调整教学。同时,受益于大数据提供的极具价值的反馈信息,教师也将身兼学习者的角色而不断进行学习。而从更广泛的意义上来说,教育不再被视为主要由教师向学生传递知识的单向过程,而成为一种将为包括学生在内的每一个人提供学习、提高和发展机会的场所。

除了目睹这些为人类学习带来强大而直接的影响外,我们还会看到大数据在整体上将会重塑教育系统,影响整个系统的决策和规划。比方说,大数据可用于合理预测在发展中的城市里设置新的学校和幼儿园的具体位置。在更广泛的层面上,大数据赋予我们重新审视事物的崭新视角,发现那些促使某类学校脱颖而出的因素。例如,我们能够不仅仅判断职业培训是否为受训者提供了恰当的职业路径,还可以掌握实施职业培训的条件及背景。

过去,在欧洲及北美地区,教育系统的架构在很大程度上依赖于当时的主要决策制定者的偏好,而不是依赖于数据。显而易见的是,这样的做法无法获得最佳成效。利用大数据,我们可以使决策者得以在全面而坚实的经验基础上改善其决策的质量,从而使教育决策从意识形态的偏见中脱离出来。

大数据还将改变教育提供者的版图。在小数据时代,绝大多数教育系统是由那些具有悠久历史和重要品牌的已有机构所控制,而新来者则缺乏这些机构所拥有的源于知识和经验的声誉。但这种状况正在发生着变化。智能手机应用公司多邻国(Duolingo)是这一方面的绝佳个案。由于大数据所提供的洞察机会,多邻国这一教育新提供者能够比许多离线同行提供更好的外语培训。基于公司所收集的人们如何学习语言的数据,多邻国自然比其同行更好地认识到如何最优化地学习一门外语。简言之,大数据将由原有的和新近的教育提供者所组成的充满生机的生态系统创设一种厚实的基础,这在学习时间和学习场所正变得无处不在和无时不在的今天表现的尤为明显。

大数据形塑教育系统,还有另外一种方式。随着我们开始认识到源自大数据的力量和价值,我们需要教授学生必要的技能来利用这种价值。正如我们在《与大数据同行》一书中所论证的,这就需要获得大数据统计和应用数学以及机器学习 and 数据碎片处理方面的专门知识;还需要深度理解大数据分析的局限性,以避免陷入数据产生歧义之陷阱。最后但同样重要的是,必须从伦理上来理解,并不是每一次大数据的利用都是合适的和被允许的。

大数据将会改变学生未来如何看待他们周围的世界并认识这一世界。学生会比以往任何时候都努力去收集并分析数据,以丰富他们做出决定时的所需信息。他们还会接受这样的观点,即世界远比我们想象的要复杂的多,但通过大数据我们可以有计划来逐步欣赏这一复杂性,尽管我们可能不会完全掌握它。

二、大数据关照下的数字鸿沟问题

赵和张:大数据的力量是如此强大,“得数据者得天下”。那么,那些数据巨头如谷歌和可汗学院(Khan Academy)与那些无法获取海量数据的机构之间的

数字鸿沟是否会愈演愈烈？未来又将会何去何从？大数据对个人学习者带来的挑战是什么？是不是会使优秀者更优秀？不会利用数据学习的人改进提高的机会更少，从而加剧数字鸿沟？

舍恩伯格和库克耶：大数据的最重要部分并不是技术或算法（algorithms），而是数据本身，其结果是谁拥有了数据或谁接触到数据谁就有强力。这至少要冒教育的数字鸿沟（digital divide）之险，但也未必一定如此，有两个原因。第一，来自社会各阶层的多数学生都会学会使用大数据，因此工具的应用会非常的广泛。第二，学校官员和监管者能够设计出政策来缓解这一分界的负面影响，如果这种影响确实发生的话。

那么教育的数字鸿沟又是如何的呢？我们可以设想一下，那些就读于名校的优秀学生极易接触到优于其他系统的大数据学习系统。我们可以作这样的预测：来自名校的位于同一智力水平和入学考试分数的学生会比那些只能应用不太优质的大数据学习系统的同伴们学得更多更快且考试分数更高，因为后者就读的学校没有财力提供优质的大数据学习系统。随着时间的推移，应用最佳大数据系统的学生在校中会表现出众，在以后的职业生涯中也会表现出众。

这是种非常糟糕的境遇，但却又很可能存在于今日社会生活的各个领域，包括教育领域。最聪慧的学生能到最优质的学校读书，他们在那里获得最多的资源得到最好的培训。富裕的父母则把自己的孩子送到实施定制教育的私立学校，而这是普通家庭的孩子所无法获得的教育。这种鸿沟已经存在。因此重要的是，我们不能寄予大数据比寄予世界其他方面以更多的期望。

当然，缓解教育数字鸿沟的潜能还是存在的。那为什么这又不是必然发生呢？因为如果所有学生都获益于大数据，那么我们正在努力改善部分群体这一事实所产生的实际效果，就会小于让所有学生都能享受这样的改善形式。考虑一下计算机和移动手机吧。当计算机和移动手机已经成为城市地区人们的主流工具时，一个人是拥有高端的在线 iPhone 还是中等特征的安卓系统智能手机就显得不那么重要了。与过去人们没有移动手机时相比，智能手机表现行为方面的差异尽管是真实存在的，但远远小于人们从中所能获得的益处。

重要的是，如果教育中的大数据看起来在造就不公平的优势和促成新的不平等，那么公共政策就要跟进了。例如，我们可以说学生拥有自己的数据，而不是大数据教育平台拥有它们。管理者可以要求“数据的便携性”，这样就会存有一种流动性市场（fluid market），让学生和学校带来自己的数据并获得服务。这就会减少大数据和教育集聚平台的力量。

今天，大数据是新事物，就像我们在 20 世纪 50、60 年代看待计算机一样，那时只有最大的公司才能使用大型且昂贵的机器来从事大型任务并聘请专家予以操作这类机器。但是在今天，计算能力存在于每个人的口袋中，或者说可以向购买水和电一样，向诸如 Amazon Web Services 此类的云计算公司购买计算能力。教育中的大数据同样如此：它并不只属于最好学校中的最优秀学生，大数据存在于任何地方。

三、大数据时代背景下的教师和学校管理者

赵和张: 你们在书中提到反馈(feedback) , 个性化(individualization) 和概率预测(probabilistic predictions) 是大数据在教育领域有效发挥作用带来的三种重要的变化。从本质上讲, 这三种变化是基于大数据的评价带来的, 虽然如你们提到, 一些政府部门希望建立机制从源头保证大数据的安全和使用的安全, 但是对使用大数据的普通教师、研究者管理者来说需要具备怎样的资质和能力?

舍恩伯格和库克耶: 随着工商企业和机构组织围绕着大数据重建其获益模式、文化和工作流程, 日渐清晰的情况是, 在先前环境下获得成功所必需的技能已经不同于今天所必需的技能。第一个变化是思维方式, 就如我们在《大数据时代》一书中所称作的“大数据思维倾向”。这一概念表明一个人需要在看待整个世界以及世界中的所有事物时, 要从物质事物转向交互作用, 并把它看作是一个收集和分析数据的平台。

如果学校这样做时, 它们会重新把自己想象成类似于工厂一样的场所, 因为工厂中的所有的投入和产出需要可以尽可能有效予以测量。但是这里, 衡量效率的尺度就是什么才能最好地帮助到学习和学生学业成绩, 而不是对管理者和教师来说最容易做的事情, 这是学校系统诸多方面过去之所存在的理由。大数据的思维模式使教师和学校官员尽其所能测量和检测所有事物, 以便发现为支持学生的进步在怎样的传导功能下做什么才最为有效。

然而, 学校管理者在这样做时需要面临第二种挑战, 因为作为一个教育家所从事工作的本质发生了变化。教师和管理者自身需要成为“数据脱盲者”(data literate)。他们需要知道如何通过阅读图表来追踪学生的进步; 他们需要知道如何通过分析概率预测, 从而使他们能够解释这对学生意味着什么, 并鼓励学生以某种方式而不是其他方式才能更有效地工作。

大数据不会破坏教学工作, 至多只是改变它们, 迫使教师成为学生成长过程中的合作伙伴。由于教学的某些基础而单调部分可以为计算机所取代——成为慕课课程, 依教学任务或分级考试顺序而确定——教师需要将更多有价值的內容带给学生, 并成为大数据学习系统与学生之间的连接点(interface)。

第三, 学校变革的本质才是最重要的。这是大数据分析所需的基础设施。因此, 学校系统突然就需要聘用那些能够熟知网络和数据存储的人员, 他们是网络安全方面的专家, 他们知道如何在建立系统以分享数据的同时注重保护隐私。而实现这一目标所需的技能对于纸质文件和小数据时代的学校而言是全新的。当教育数据未被用来使教学适应个别的学生, 这些职业特性就没有成为学校管理的特征。但到明天它们就会是, 因为其目的在于使学校支持为世界培养学生的学校使命, 并与此使命相适应。

四、大数据与求变且渐变的学校教育模式

赵和张: 《与大数据同行》书中提到可汗学院, MOOCs(慕课) 和 School of

One 这样的新教育模式,它们显然都特别依赖于从在线教育平台上获取数据。但是实际上,学校教育中的面授教育依然占据重要位置。您认为未来的教育模式从整体上说,是不是在线教育在正式教育中所占的比重也会越来越大?如果是这样,那学校教育面临的挑战是什么?

舍恩伯格和库克耶:在今后的教学过程中,教师和学校的确需要增加更多的价值。慕课虽不会颠覆现有的教育模式,但它会渗透进所有学生必修的基础课程中。尽管如此,大多数教师仍然不愿意教授慕课课程,因此只能由教育主管指派资历最浅的新教师去任教。然而,如果负责这些基础课程的是一位杰出教授,则能够更多地启发和激励学生,并通过增强他们对学科的爱来改变其人生。

最理想的情况是,这些基础课程均由世界上最优秀的专家来教授,而不是随便碰到谁就由谁来教。慕课在大学的导论性课程中和高中层面的基础教学中将大展身手,首先将是自然科学和数学,因为在这两门课程中解释和局部敏感问题不那么重要。但慕课一定会找到恰当的方式融入所有学科领域中。

然而,学校依然会保持面授教学的传统。但既然其他学校都可以接入同一个慕课课程网络,而且这些精品基础课程将出现在绝大多数的教育机构,那我们学校也必须这么做。在普通课堂层面不会存在什么对立的情况。由于基础课程都是相同的,因此重点的区别就是面授的课堂体验了。

在慕课时代,学校需要比以往将更多的思考、关注和财力投入到面授的课堂教学中去。这与二十世纪70年代到90年代之间,医学在西方出现的转变很相似。那时医生的角色在从单一的诊断病情和告知患者病情逐步向人性关怀转变。医生不仅需要向患者解释医疗方案背后的原因,还需要跟踪观察患者经历的每一步诊疗过程。患者需要知情,并能得到权威的答案。

学生与教师之间的互动可以遵循同样的方式。虽然传统的师生等级制度仍将存在,但它的严格性将逐渐降低。教师和学生会发现,他们在学习过程中的角色更像是合作者。与教师在一起的时间会更有价值,且更为有效,因为学生可能会得到围绕着概念和方法而量身定制的个性化教学以提高自己的学业成绩,从而超越大数据系统提出的要求。

学校将要面临的挑战是要挑选、培养、管理并留住那些具备作为一个教练所拥有之个人技能的教师,而不只是仅仅掌握其所教课程之知识的教师。这就需要重新设想未来的课堂设置,并重新设计与之匹配的课程。课堂中的设置需要改变。如果教学是个性化和在线的,而不需要教师同时向所有学生教授相同的内容,那还有什么必要一定要学生们排排坐听课呢。

也许常见的工作台,才是最合适的课堂配置。毕竟,学生最终要加入公司,而那里的知识型工作者将会根据在给定时间内所要完成的项目而使用“公用办公桌”(hot desk)。大数据已经改变了教和学的本质,难道这种意识不应该先从学校开始培养吗?

五、大数据的潜在威胁与可能的应对策略

赵和张:《与大数据同行》还提到大数据在为教育带来巨大变革的同时,也伴随着一些潜在的威胁,如永久的过去(permanent of the past)、规定了的未来(fixed future)等,你们认为这些威胁是否会成为大数据演进的障碍?应该如何来消除这些负面的影响?

舍恩伯格和库克耶:在教育领域,不加选择的随意使用大数据的确会产生重大风险。举例来说,如果教育机构使用大数据来预测学生个体的职业生涯,然后迫使他们接受这样的职业发展轨迹,那么父母理所当然地会担心他们的孩子被剥夺了未来。在今天这样的世界里,人们可能不再愿意让自己的个人数据被收集被使用,因为这些数据有可能被用来阻止人们去学习他们热衷的专业,并不是用于改善学习的过程。同大数据相比,在如今这一普遍失信的世界里,数据不仅将越来越难收集,而且人们还可能会提供虚假的数据来“玩弄数据系统”。其结果将是今后依此作出的决策会比现在更糟糕。

在教育领域运用大数据,至关重要的一定是在互信的环境中得到学生和家长的信任。目前西方的隐私法案在灌输这种信任方面并不是特别的有效。因此,例如在美国有家长抗议数据平台收集和分析学生的敏感信息,就可以理解了。为了让大数据在教育领域发挥它的潜力,人们就要求新的隐私法案能够规范个人敏感信息的使用,而不是纠结于学生(或家长)是否同意数据被收集。

教育领域隐私规则对于如何使用大数据的新关注,将要求教育组织认识到道德的挑战并伴以极其谨慎的行为。这也同样要求监管者确保严格执法。只有当这些要素全部具备,学生和教师才能够相信大数据将会促进他们的学习并帮助他们实现梦想,而不是妨碍他们。

此外,我们可能需要一种新的大数据专家。我们在《与大数据同行》一书中称之为算法专家(algorithmists)。这些算法专家将在大数据分析方面进行专门训练。如果学生和家长察觉大数据预测出现差错并开始危害到自己,他们就会及时求助于这些算法专家。算法专家必须像医生或土木工程师那样恪守职业道德,包括公正和保密。他们可以要求教育机构授权他们访问数据和算法以验证分析的有效性,这样就可以透视大数据这一在教育预测上的黑匣子。

随着时间的推移,通过不断曝光大数据应用在诸如教育这样的敏感领域中出现的问题,算法专家不仅可以进一步促进大数据会得到正确且恰当应用的公众信任,而且还能通过剔除教育系统中的“害群之马”和促进最佳实践的创设,来推动大数据在教育系统中的广泛应用。

六、大数据背后其实是人的问题

赵和张:最后再问一个问题,本书(E-book)上市已有数月,至今有无新的

动态可以与中国读者分享?

舍恩伯格和库克耶: 在过去几个月里,《与大数据同行》这本书的问世,如我们预期的那样获得了很大进展。那些已经在使用大数据推动学习的个人或机构如吴恩达(Andrew Ng)和可汗学院仍在继续这么做。多邻国网站在有效地帮助全球数百万人学习外语,网站甚至在服务中又增加了新的语言组件。越来越多的新创公司正在利用源自大数据分析的商业观察而进入教育市场。

也许是因为畏惧,西方的传统教育机构对于大数据带来的机遇反应冷淡。教育界的掌门人仍滞足于确立一些大数据计划,而不去开创性地思考如何利用大数据来促进教学和学习,并积极应对来自私营的新创公司不断加剧的竞争。

尝试推动现行教育系统及身处其中的学生和家长进入大数据时代的具体实践,往往会遭受重大挫折。卡内基基金和盖茨基金会资助成立了名为 inBloom 这一旨在收集和存储教育数据的非盈利机构,但在遭到家长和公众因担忧数据隐私问题的强烈抗议后, inBloom 只好取消了它在美国三个州准备实施的计划。

这个例子充分说明,任何人要在教育规划中应用大数据,取得学生和家长对其教育数据可被分析的信任是至关重要的。就像 inBloom 的失败所显现的,过度复杂的设计会让实施过程显得不够透明,让人难以理解。另外,若无法体现大数据明显的优势,应用大数据也就很难取信于人。

上述案例给我们上了很好的一课: 现行的教育机构,若不特别重视学生和家长对于数据滥用和隐私的关注,那么它们在应用大数据方面会面临很大的困难。令人啼笑皆非的,正是那些有可能从应用大数据中获利颇丰的机构,同时也将在应用大数据过程中面临最大的麻烦。

赵和张: 非常感谢你们对我们所提问题的回答和阐述,同时衷心希望《与大数据同行》一书能够受到社会各界,尤其是教育界的广泛关注。

Learning and Education with BIG DATA

Interview with Professor Viktor Mayor – Schonberger and Mr. Kenneth Cukier

ZHAO Zhongjian & ZHANG Yannan

(Institute of Curriculum and Instruction, East China Normal University, Shanghai 20062, China)

Abstract: Questions of the interview are mainly focused on the following aspects: big data and re-shaping of education system, digital divide or gap under big data, new requirements for teachers and educational administrators in the era of big data, big data and the changing education models, the potential threats of big data and the possible corresponding strategies, and the issue of human being behind big data. Professor Viktor and Mr. Kenneth make detailed responses to the above aspects.

Key words: big data; learning; MOOCs; privacy

(责任校对: 董泽华)